

Contributed Paper

## A quick-start guide to BOND: Bayesian Oxygen and Nitrogen abundance Determinations in H II regions using strong and semistrong lines

N. Vale Asari<sup>1</sup>, G. Stasińska<sup>2</sup>, C. Morisset<sup>3</sup>, and R. Cid Fernandes<sup>1</sup>

<sup>1</sup>*Departamento de Física-CFM, Universidade Federal de Santa Catarina, C.P. 476, 88040-900, Florianópolis, SC, Brazil*

<sup>2</sup>*LUTH, Observatoire de Paris, CNRS, Université Paris Diderot; Place Jules Janssen 92190 Meudon, France*

<sup>3</sup>*Instituto de Astronomía, Universidad Nacional Autónoma de México, Apdo. Postal 70264, México D.F., 04510 México*

**Abstract.** We present a quick-start guide to BOND, a statistical method to derive oxygen and nitrogen abundances in H II regions. BOND compares a set of carefully selected strong and semistrong emission lines to a grid photoionization models. The first novelty, in comparison to other statistical methods, is that BOND relies on the [Ar III]/[Ne III] emission line ratio to break the oxygen abundance bimodality. In doing so, we can measure oxygen and nitrogen abundances without assuming any a priori relation between N/O and O/H. The second novelty is that BOND takes into account changes in the hardness of the ionizing radiation field, which can come about due to the ageing of H II regions or the stochastically sampling of the IMF. We use the emission line ratio He I/H $\beta$ , in addition to commonly used strong lines, to constrain the hardness of the ionizing radiation field. Finally, we also stress the pragmatic considerations behind our Bayesian inference.

### 1. Why a statistical method based on photoionization models

When direct temperature measurements are missing, statistical methods are used to infer abundances in giant H II regions. There are two families of statistical methods. One is based on calibrating samples of objects for which the abundance could be derived from temperature-based methods. The other is based on photoionization model grids. The latter is free from observational biases, but the grid must cover all the configurations that could be found in nature. BOND (Bayesian Oxygen and Nitrogen abundance Determinations) belongs to the second family. BOND infers oxygen and nitrogen abundances using carefully selected strong and semistrong lines by comparing them to a grid of photoionization models. The source code is open and freely available at <http://bond.ufsc.br/>. Full details can be found at Vale Asari et al. (2016). This manuscript is intended to be a quick-start guide to highlight the most important aspects of the method.

## 2. What sets BOND apart

Common strong line methods based on simple calibrations ( $[\text{O III}]/[\text{N II}]$ ,  $[\text{N II}]/\text{H}\alpha$ ,  $([\text{O II}] + [\text{O III}])/\text{H}\beta$ ) assume that emission-line nebulae are a one-parameter family, and that such parameter is the oxygen abundance. The first to realise the importance of introducing a secondary parameter to measure abundances was McGaugh (1991), who considered the effect of the ionization parameter  $U$ . Nowadays there is a plethora of methods to measure abundances based on the comparison of observed to theoretical emission lines from a grid of photoionization models (McGaugh 1991; Kewley & Dopita 2002; Tremonti et al. 2004; Dopita et al. 2013; Pérez-Montero 2014; Blanc et al. 2015). We will discuss the difference of our method with respect to others.

One novelty in our grid of photoionization models is that we do not impose any *a priori* relation between N/O and O/H. The only other method that also does not tie in the nitrogen and oxygen abundances is the one by Pérez-Montero (2014), with the difference that his method uses auroral lines (so it is not a strong line method), and it selects models around the empirical N/O versus O/H relation if auroral lines are not available.

Another novelty in our method is that it uses Bayesian inference to measure abundances, not quite unlike the one by Blanc et al. (2015). The curious reader is referred to Section 4 for the key points; we warn that, although Bayesian inference is part and parcel of our method, it is not the most important aspect of BOND.

The killer features in BOND are: (a) N/O is free to vary, (b) it distinguishes between the lower and upper metallicity branches, and (c) it considers the effect of varying the hardness of the ionizing radiation field. The next section shows how we have tackled (a), (b) and (c), and explains the reasoning behind choosing which emission lines need to be fitted.

## 3. Input emission lines for BOND

Our grid of photoionization models spans a wide range in O/H, N/O, and  $U$ . The ionizing radiation field is provided by the instantaneous starburst models of Mollá et al. (2009) for six different ages. Two nebular geometries are considered (thin shell and filled sphere). We thus have five parameters in our models. Even though we are interested in measuring only two of them, O/H and N/O, we still ought to have good constraints for the other three ‘uninteresting’ parameters:  $U$ , the hardness of the ionizing radiation field, and the density structure. That is why we have chosen a set of emission lines carefully tailored to constrain those five parameters all at once.

In the following we list these emission lines and explain the reason behind each of them. Note that the (reddening-corrected) intensities of *all* of those lines with respect to  $\text{H}\beta$  are needed to run BOND.

- The strong lines  $\text{H}\beta$ ,  $[\text{O II}]\lambda 3726 + \lambda 3729$ ,  $[\text{O III}]\lambda 5007$ , and  $[\text{N II}]\lambda 6584$ . They were chosen because, to first order, the strong line ratios  $([\text{O III}] + [\text{O II}])/\text{H}\beta$ ,  $[\text{N II}]/[\text{O II}]$ , and  $[\text{O III}]/[\text{O II}]$  map into O/H, N/O and  $U$ , respectively. We fit line intensities with respect to  $\text{H}\beta$ , and not the latter

strong line ratios directly, because we assume that the intensities with respect to  $H\beta$  follow a Gaussian distribution when we calculate likelihood probabilities.

- The semistrong lines  $[\text{Ar III}]\lambda 7135$  and  $[\text{Ne III}]\lambda 3869$  (plus upper limits for auroral lines of  $\text{O}^{++}$  and  $\text{N}^+$ ). The ratio  $([\text{O III}] + [\text{O II}])/H\beta$  is bi-valued with respect to the oxygen abundance (see the inverted U shape of the relation in Fig. 1, left). That is why some methods impose a fixed relation between N/O versus O/H in their photoionization models (e.g. Dopita et al. 2013; Blanc et al. 2015). Since we are interested in inferring both N/O and O/H, we did not want to use nitrogen lines to break the bimodality in the oxygen abundance. Restricting our search to emission lines easy to measure in typical optical spectra (though not always reported in the literature), we found an ideal candidate in the emission line ratio  $[\text{Ar III}]/[\text{Ne III}]$ .  $\text{Ar}^{++}$  and  $\text{Ne}^{++}$  are formed in roughly the same zone, but the excitation potentials of  $[\text{Ar III}]$  and  $[\text{Ne III}]$  are very different (1.7 and 3.2 eV, respectively), so the ratio of these lines is sensitive to the electronic temperature. Argon and neon are primary elements and their global abundance ratio is expected to be constant. Besides, they are both inert, thus do not suffer dust depletion, so their abundance ratio in the ionized gas phase remains constant. Fig. 1 (centre) shows the line ratio  $[\text{Ar III}]/[\text{Ne III}]$  in our grid as a function of  $[\text{O III}]/[\text{O II}]$  (the latter traces the ionization parameter  $U$ ). The points are colour-coded as falling in the lower or upper metallicity branch (blue and red, respectively), showing that  $[\text{Ar III}]/[\text{Ne III}]$  can break the bimodality in the oxygen abundance. An extra help on finding the right metallicity branch can come from *upper limits* in auroral lines.
- The semistrong line  $\text{He I}\lambda 5876$ . This is the crucial part of BOND. Since we consider different ionization scenarios, e.g. different spectral energy distributions (SED) of the ionizing radiation, we need to infer which of those scenarios is more appropriate. We expect the SED to be different in different H II regions due to ageing (so the most massive stars have disappeared) or due to stochastic effects in low luminosity H II regions when the upper part in the stellar initial mass function of the ionizing cluster is not fully sampled. Fig. 1 (left) shows how the emissivity of  $([\text{O III}] + [\text{O II}])/H\beta$  depends not only on the oxygen abundance (abscissa), but also on the hardness of the ionizing radiation field (colour code). Note that at high metallicities a single value of  $([\text{O III}] + [\text{O II}])/H\beta$  can span 1 dex in oxygen abundance for different ionizing radiation fields. Fig. 1 (right) shows that the  $\text{He I}\lambda 5876/H\beta$  line can be used as a proxy of the hardness of the ionizing radiation field.

#### 4. Why go Bayesian

First, let us emphasise that the Bayesian inference is *not* what sets BOND apart. The heart and soul of BOND is the set of carefully selected emission line ratios, tailor-made both to infer the oxygen and nitrogen abundances in H II regions, and to take into account important secondary parameters. As reasoned in the

previous section, this allows us to (a) break the bimodality in oxygen abundance without any a priori relation between N/O versus O/H, and (b) consider the (previously neglected) role of the ionizing radiation field in abundance determinations.

That said, we have opted for Bayesian inference for good reasons. Before diving into the most important Bayesian aspects, let us examine the method’s acronym. BOND stands for Bayesian Oxygen and Nitrogen abundance Determinations. Our method is not called Oxygen and Nitrogen abundance, Ionization parameter, and Hardness of the ionizing radiation field Determinations (ONIHD). The reason why the letters IH are not in the method’s name is the same reason why we have decided to introduce the B for Bayesian in its name: the Ionization parameter and Hardness of the ionizing radiation field are *nuisance* parameters, and the only way to get rid of them respecting dimensional analysis is by going Bayesian.

Let us lay out the problem to see how its resolution points to Bayesian inference. We start with a carefully designed photoionization grid, finely spaced and spanning a wide range in O/H, N/O, and  $U$ . It also considers a few values of hardness of the ionizing radiation field mimicked by different SED ages. Even though the latter two parameters ( $U$  and ionizing radiation field) are important and need to be well modelled, they are of *secondary* interest. This is why our acronym shifts from ONIHD to OND.

How do we get rid of parameters for which we do not care (a.k.a. *nuisance* parameters)? If we want to consider the probabilities of all models in our grid at the same time, then we can simply *marginalise* over the nuisance parameters. Marginalising is nothing more than integrating over a parameter. For instance, for a fixed O/H and N/O, we marginalise over  $U$  simply by adding up all the probabilities of all models of a given O/H and N/O for all  $U$ . The trick is that to integrate, say, in  $dU$ , the probability density function (PDF) in the integrand must have physical units of  $U^{-1}$ . Ordinary likelihood PDFs (e.g.  $e^{-0.5\chi^2}$  for Gaussian distributions<sup>1</sup>) have units which are the inverse of the observational data being fitted. So, from a *pragmatic* point of view, we are obliged to write out the posterior PDFs, which have the correct physical units when we integrate over a model parameter. For a thorough argument on the dimensional analysis of PDFs, see Hogg (2012), especially the discussion around his equation 3.

In other words, going Bayesian gives us licence to kill the nuisance parameters, so we need to add B to OND. To write the posterior PDFs, we need to spell out our priors. This has a two-fold benefit. First, we can plug in an informative prior: if we have empirical evidence that some models are more probable in nature than others, we can give them more weight by setting the prior probabilities just right. In our code so far, we have taken the most conservative approach we can and, following Blanc et al. (2015), we use an uninformative prior (specifically, a Jeffrey’s prior that is logarithmic in O/H, N/O and  $U$ ).

The second benefit of setting a prior is that we have an explicit prescription for making a finer grid. The problem of comparing data to a uniformly spaced grid of models by using a  $\chi^2$  likelihood is that most models will be very distant from the observed data as measured by the uncertainties  $\sigma_j$ . If a grid is very

---

<sup>1</sup>  $\chi^2 = \sum_j (c_j - o_j)^2 / \sigma_j^2$ , where  $o_j$  and  $\sigma_j$  are the observed line intensity and its uncertainty, and  $c_j$  is the computed line intensity.

very rough, the closest model might even be a few  $\sigma_j$  away from the model with the highest likelihood. An ad hoc prescription to deal with this problem is by setting up cooking factors to increase the observed uncertainties, thus decreasing the distances between observed and computed emission lines. A cooking factor has no real justification and needs to be tailored to work for each new data point. Since we are using a Bayesian prescription, we can do much better than relying on ad hoc prescriptions. We simply interpolate our grid where we need—and interpolation in  $\log N/O$ ,  $O/H$  and  $U$  is reasonable once the grid is fine enough. The interpolation is informed by posterior probability of each element in the grid: if an element has a high probability, it is worth creating more grid points inside its volume<sup>2</sup>.

Just a final note on the Bayesian parlance. The outcome of a Bayesian inference is the posterior PDF for a set of model parameters (say,  $O/H$ ,  $N/O$ ,  $U$ , hardness of the ionizing radiation field). Since we are interested only in  $O/H$  and  $N/O$ , we can integrate out all other parameters and obtain the *joint*  $O/H$  and  $N/O$  posterior PDF. The joint PDF is simply a two-dimensional function that gives the probability of each point in the  $O/H$  versus  $N/O$  plane, which is the ultimate goal of BOND.

However, sometimes it is impractical to work with the full joint PDF, so we need a summarised description in the form  $12 + \log O/H = 8.35 \pm 0.02$ . There are many ways to transform a two-dimensional function into a nominal value and a dispersion. One way is to set the nominal value for  $N/O$  and  $O/H$  to be the point where the joint PDF is the highest. We call this number the maximum a posteriori (MAP), because it is calculated after (i.e. a posteriori) the marginalisation of nuisance parameters. For the dispersion, we can define ellipses of credible regions that encompass, say, 5, 50, 68 or 95% of the total joint PDF.

If we want to marginalise away either  $O/H$  or  $N/O$  (i.e. if we are interested in one of those parameters alone), we can integrate over the other parameter. Summarising the fully marginalised PDF also opens up a menu of choices. The nominal value can be taken as the mean, median or mode (that is, its peak) of the PDF. For the dispersion, the usual choices are either the 50, 68, or 95 percent equal-tailed or highest density intervals. Note that the equal-tailed intervals are related to the median. The median is the point in the PDF curve where 50 percent of the probability is to the left and 50 to the right. The equal-tailed 68% interval is the region of a curve where 16 percent of the probability is to the right and 16 percent to the left. The highest density intervals, on the other hand, are related to the mode. The mode is the point in the curve of highest probability. The 68% highest density interval is the region around the mode that adds up to 68 percent of the total probability. A nice visualisation tool for those descriptions can be found at <http://www.sumsar.net/blog/2014/10/probable-points-and-credible-intervals-part-one/>.

---

<sup>2</sup>The algorithm to do importance sampling in BOND is the octree sampling, which is computationally inexpensive. The reader might be more familiar with MCMC samplers, which are more adequate when one has to compute models on the fly and does not have a pre-defined grid. In our case, it is much more sensible to compute many photoionization models a priori and interpolate them on the fly than generating photoionization models on the fly.

Another minor nuisance parameter we marginalise away are the uncertainties. For  $[\text{Ar III}]/\text{H}\beta$ ,  $[\text{Ne III}]/\text{H}\beta$  and  $\text{He I}/\text{H}\beta$ , we consider an extra noise source added in quadrature to the observational uncertainties which we allow to vary from 2 to 100 percent of the line intensity. The extra noise source is needed because, in nature, the Ar/O and Ne/O ratios may differ somewhat from the ones assumed in our model grid. Regarding  $\text{He I}/\text{H}\beta$ , the problem is that our grid is only coarsely meshed as regards the hardness of the ionizing radiation field. We then calculate the marginalised likelihood PDF for those lines by using all values of this extra noise and then marginalising it away. We do so because we do not expect those lines to be completely correct in our photoionization models, and they are used only to infer secondary parameters.

## 5. Conclusions

We have highlighted the main characteristics of BOND, a method based on a grid of photoionization models to measure oxygen and nitrogen abundances in giant H II regions using strong and semistrong lines. We show why it is important to consider secondary parameters in abundance determinations, especially the hardness of the ionizing radiation field. The SEDs in H II regions can vary from region to region due to stellar ageing or to the stochastically sampling of the IMF. We also show how one can break the metallicity bimodality without recourse either to auroral lines or a fixed relation between N/O and O/H. We use a selective set of emission lines to infer all those parameters. Finally, we argue why using Bayesian inference is the correct way (motivated by the dimensional analysis of probability density functions) to treat the secondary parameters in abundance determinations.

## 6. Acknowledgements

NVA is grateful to Oli Dors for having organised such a useful workshop. NVA acknowledges the support from Programa de Pós-Graduação em Física da UFSC and CAPES/PROAP to attend the workshop. GS and NVA acknowledge the support from the CAPES CsF-PVE project 88881.068116/2014-01. The grid of models has been run on computers from the CONACyT/CB2010:153985, UNAM-PAPIIT-IN107215 and UNAM Posgrado de Astrofísica projects.

## References

- Blanc, G. A., Kewley, L., Vogt, F. P. A., & Dopita, M. A. 2015, *ApJ*, 798, 99
- Dopita, M. A., Sutherland, R. S., Nicholls, D. C., Kewley, L. J., & Vogt, F. P. A. 2013, *ApJS*, 208, 10
- Hogg, D. W. 2012, *ArXiv e-prints*
- Kewley, L. J. & Dopita, M. A. 2002, *ApJS*, 142, 35
- McGaugh, S. S. 1991, *ApJ*, 380, 140
- Mollá M., García-Vargas M. L., Bressan A., 2009, *MNRAS*, 398, 451
- Pérez-Montero, E. 2014, *MNRAS*, 441, 2663

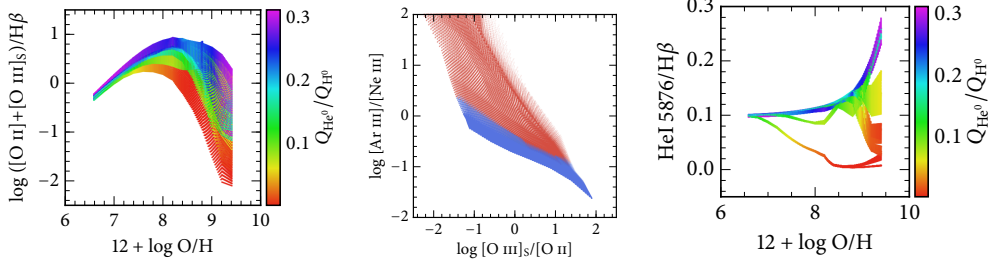


Figure 1. **Left:**  $([\text{O III}]_{\text{S}} + [\text{O II}])/\text{H}\beta$  versus  $\text{O}/\text{H}$  coloured by  $Q(\text{He}^0)/Q(\text{H}^0)$ , which traces the hardness of the ionizing radiation field. This represents the two secondary effects considered in BOND. First, for a given  $Q(\text{He}^0)/Q(\text{H}^0)$ ,  $([\text{O III}]_{\text{S}} + [\text{O II}])/\text{H}\beta$  maps into two different  $\text{O}/\text{H}$  values. We find the correct metallicity branch by using the  $[\text{Ar III}]/[\text{Ne III}]$  ratio (centre). Second, for high metallicities  $([\text{O III}]_{\text{S}} + [\text{O II}])/\text{H}\beta$  span almost a decade in  $\text{O}/\text{H}$ . We use  $\text{He I } 5876/\text{H}\beta$  (right) to find the correct hardness of the ionizing radiation field. **Centre:**  $[\text{Ar III}]/[\text{Ne III}]$  versus  $[\text{O III}]_{\text{S}}/[\text{O II}]$ , blue for models in the lower metallicity and red for models in the upper metallicity branch. **Right:**  $\text{He I } 5876/\text{H}\beta$  versus  $\text{O}/\text{H}$  coloured by  $Q(\text{He}^0)/Q(\text{H}^0)$ .  $\text{He I } 5876/\text{H}\beta$  can be used as a proxy for  $Q(\text{He}^0)/Q(\text{H}^0)$ , except for the highest values of  $Q(\text{He}^0)/Q(\text{H}^0)$ . All figures are based on our grid of photoionization models and taken from Vale Asari et al. (2016).

Tremonti, C. A., Heckman, T. M., Kauffmann, G., Brinchmann, J., Charlot, S., White, S. D. M., Seibert, M., Peng, E. W., Schlegel, D. J., Uomoto, A., Fukugita, M., & Brinkmann, J. 2004, ApJ, 613, 898

Vale Asari, N., Stasińska, G., Morisset, C., & Cid Fernandes, R. 2016, MNRAS, 460, 1739